

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-224692

(43)Date of publication of application : 03.09.1993

(51)Int.Cl.

G10L 3/00

(21)Application number : 04-027019

(71)Applicant : NEC CORP

(22)Date of filing : 14.02.1992

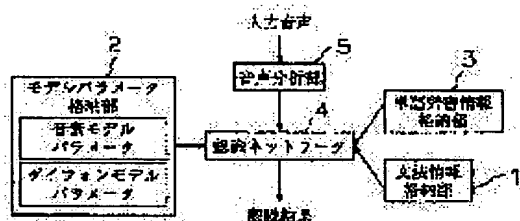
(72)Inventor : ISOTANI RYOSUKE

(54) CONTINUOUS SPEECH RECOGNITION SYSTEM

(57)Abstract:

PURPOSE: To precisely recognize even a speech generated by continuously voicing words, without increasing the throughput at word borders when recognition units depending upon environment.

CONSTITUTION: Diphones obtained by fractionizing a phoneme by a following phoneme and a phoneme which does not depend upon the following phoneme are used as the recognition units. A word dictionary 3 is so constituted as to use a phoneme model at the tail of a word and a diphone model in other cases. A recognition network 4 is generated by using the word dictionary, model parameters, and grammatical information to recognize a continuous speech. The parameters of the phoneme model are found by averaging the parameters of the diphone model.



LEGAL STATUS

[Date of request for examination] 30.07.1996

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3033322

[Date of registration] 18.02.2000

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(3)

同様、前後の単語に応じて環境に依存した認識単位を用いているので、性能の低下はない。また、語頭、語尾といった単語内の位置を考慮なくともよいので、認識単位を環境の長さの長いサブワードのパラメータとは独立にモデル（あるいは隠層パターン）のパラメータを学習データから求める場合でも、文音声のデータは必ずしも必要としない。しかし一般にある単語と接続しうる単語は複数存在するので、語頭あるいは語尾にそのそれぞれ接続する単語に応じて認識単位をすべて用意しなければならない。認識時の処理量が増大するという欠点がある。

【0008】本発明の目的は、このような欠点を除去した連続音声認識方式を提供することにある。

【0009】問題を解決するための手段）第1の発明は、隣接する言葉の並びに依存して決定されるサブワードを認識単位とし、辞彙中の各単語をサブワードの列のないネットワークで記述したサブワード表記辞書と単語の接続の規約を記述した文法として、環境の長さの長いサブワードを用意しておき、環境中の長さの長い単語内に存在する環境のみに依存するサブワードを用いることを特徴とする。

【0010】第2の発明は、第1の発明の連続音声認識方式において、環境の一部分が共通であるサブワードのモデルのパラメータを平均化した値をサブワードのモデルのパラメータとして用いることを特徴とする。

【0011】第3の発明は、第1の発明の連続音声認識方式において、各サブワードを隠れマルコフモデルでモデル化し、環境の一部分が共通であるサブワードについてモデルのパラメータが等しくなるという内蔵条件のもとで、フォワード・バックワードアルゴリズムによる学習を行うことにより得られたパラメータをサブワードのモデルのパラメータとして用いることを特徴とする。

【0012】第4の発明は、第3の発明の連続音声認識方式において、フォワード・バックワードアルゴリズムによる学習を1ループだけに限定して得られたパラメータをサブワードのモデルのパラメータとして用いることを特徴とする。

【0013】

【作用】第1の発明によれば、環境の長さの長いサブワードを使用する場合、単語内だけでは環境が決まらない語頭や語尾については、単語内だけで決まるより短い環境だけに依存するサブワードを使用する。これによって、接続する単語によらない処理が可能となり、文獻3のような処理量の増大を防ぐことができる。さらに、このサブワードはその外側の環境に依存しないものであるため、たとくに語頭、語尾といった特殊な環境に限定してない。したがって、単語間が単語内と同様に連続的に発生される場合にも対処できる。また、認識単位のモデル（あるいは隠層パターン）のパラメータを学習データから求める場合でも、文音声のデータは必ずしも必要とせ

ず、単語内にあらわれないサブワードのデータだけから学習することもできる。

【0014】環境の長さの長いサブワードのパラメータとは独立に環境の長さの長いサブワードのパラメータとは独立に求めることもできるが、本方式のように環境の長さの長いサブワードと環境の長さの短いサブワードを併用して使用する場合には、環境の長さの長いサブワードが隣合って現れる場所とこれらが時間的に連続していることが必要である。たとえば、セグメンテーションを働に与えない。第2の発明によれば、環境の短いサブワードのパラメータを独立に求めると、この条件が満たされる保証はない。第2の発明によれば、環境の短いサブワードのパラメータをその環境を含む環境の長さの長いサブワードのパラメータを平均して求めるので、環境の長さの異なるサブワードの間で時間的な連続性が保たれる。

【0015】サブワードを隠れマルコフモデルでモデル化する場合、第2の発明におけるパラメータ平均化の方法は自明ではない。第3の発明では、フォワード・バックワードアルゴリズムにおけるパラメータの「結び」の手法を利用して、パラメータの平均化と同様のことが行える。パラメータの「結び」については、刊行物「標準モデルによる音声認識」中川一著に詳しく説明されている。

【0016】第4の発明では、さらに学習のループを1回だけに制限している。学習の1回目では学習データのセグメンテーションが初期モデルによって行われるので、初期モデルとして環境の長さの長いサブワードのモデルを用いて学習を1回だけに制限することにより、上記の時間的連続性の条件も満たされる。

【0017】

【実施例】本発明の実施例について説明する。認識単位として、言葉を後続の1言葉に依存して分類した「ダイフォニ」と前後の言葉に依存しない（すなわち環境の長さ0）言葉そのものを用いる例について説明する。

【0018】図1は、第1の発明の認識方式を実施する装置の全体のプロック図である。文法情報格納部1では、受入する文を有状態ネットワークで表している。

モデルパラメータ格納部2には各ダイフォニおよび音声モデルのパラメータが格納され、単語辞書情報格納部3には辞彙中の各単語の発音記号がダイフォニと音素からなる列で記述されている。文法、辞書、モデルパラメータの情報から、あらかじめ単一の認識ネットワーク4を作成しておく。認識時には、入力音声は音声分析部5で特徴パラメータの計算のために変換され、認識ネットワーク4の時間と照合が行われて、認識結果として出力される。認識ネットワーク4の作成の方法および特徴パラメータの時系列と認識ネットワークとの間の照合の方法は、文獻3に詳しく述べられている。

【0019】図2に本実施例における単語辞書情報の一例を示す。単語“cook”が音素列で“k”“u”“h”

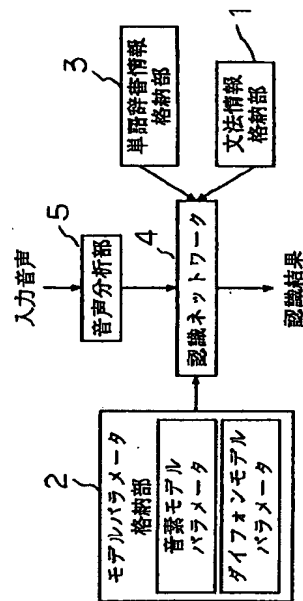
(4)

“k”と表されたとすると、ダイフォニ表記では最初の2音素はそれぞれ“k”“u”“h”と表される。ここで、たとえば“k”“u”“h”は後続音素が“u”“h”である音素“k”を表す。第3の音素については、単語内の情報だけではこのようにダイフォニで表すことができないので、後続の音素に依存しない音素“k”“*”を用いて、モデルパラメータ格納部には“k”“ax”、“k”“ih”、“...”、“s”“a”“x”、“s”“ih”、“...”などのダイフォニのモデルのパラメータに加え、“k”“*”、“s”“*”、“...”などの後続の音素に依存しない音素のモデルのパラメータも格納されている。

【0020】図3に第2の発明の実施例における音素モデルのパラメータの決定法を示す。認識方式を実施する装置の全体のプロック図は第1の発明と同じである。後続音素に依存しない音素のモデルのパラメータをダイフォニのモデルのパラメータを平均化することによって求める。たとえば、“k”“*”のモデルのパラメータは平均することによって求める。平均する際に、各ダイフォニのあらかじめ学習データなどから求めた出現頻度に応じて重みづけすることも可能である。

【0021】図4に第3の発明の実施例における音素モデルのパラメータの決定法を示す。認識方式を実施する装置の全体のプロック図は第1の発明と同じである。ダイフォニおよび音素は隠れマルコフモデル（HMM）でモデル化する。ダイフォニのHMMモデルのパラメータは通常の学習法で求める。音素のHMMモデルのパラメータは、その音素を後続の音素で分類したダイフォニのHMMモデルの対応するパラメータがすべて等しくなるような拘束条件のもとでフォワード・バックワードアルゴリズムで学習する。すなわち、“k”“ax”、“k”“ih”、“...”、“s”“a”“x”、“s”“ih”、“...”

【図1】

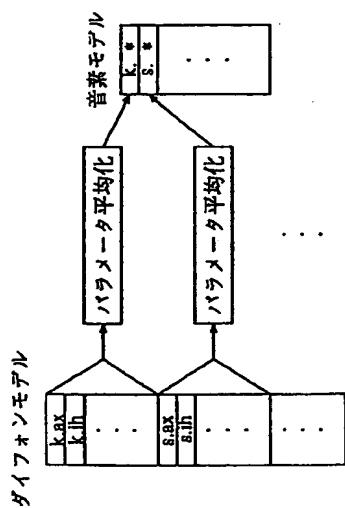


【図2】

単語	サブワード表記
...	...
cook	k u h k
cool	k u w u l
...	...

(5)

【図3】



【図4】

